A Second Look at Password Composition Policies in the Wild: Comparing Samples from 2010 and 2016

Peter Mayer and Jan Kirchner, *Technische Universität Darmstadt;*
Melanie Volkamer, *Technische Universität Darmstadt, Karlstad University*

https://www.usenix.org/conference/soups2017/technical-sessions/presentation/mayer

This paper is included in the Proceedings of the
Thirteenth Symposium on Usable Privacy and Security (SOUPS 2017).
July 12–14, 2017 • Santa Clara, CA, USA
ISBN 978-1-931971-39-3

A Second Look at Password Composition Policies in the Wild: Comparing Samples from 2010 and 2016

Peter Mayer*, Jan Kirchner†, Melanie Volkamer*‡
* SECUSO - Security, Usability, Society, Technische Universität Darmstadt
† Institute of Psychology, Technische Universität Darmstadt
‡ Privacy and Security Research Group, Karlstad University
{peter.mayer,melanie.volkamer}@secuso.org

ABSTRACT

In this paper we present a replication and extension of the study performed by Florêncio and Herley published at SOUPS 2010. They investigated a sample of US websites, examining different website features' effects on the strength of the website's password composition policy (PCP). Using the same methodology as in the original study, we re-investigated the same US websites to identify differences over time. We then extended the initial study by investigating a corresponding sample of German websites in order to identify differences across countries. Our findings indicate that while the website features mostly retain their predicting power for the US sample, only one feature affecting PCP strength translates to the German sample: whether users can choose among multiple alternative websites providing the same service. Moreover, German websites generally use weaker PCPs and, in particular, PCPs of German banking websites stand out for having generally low strength PCPs.

1. INTRODUCTION

Creating a password usually requires adherence to a password composition policy (PCP). Florêncio and Herley [5] analysed the PCPs of 75 different websites in 2010 and reported a high diversity. They investigated several website features (e.g. whether the user name is publicly visible, the value of the resources protected by the password, or whether the website advertises on other websites) in order to isolate those features that influence the PCPs' strength. They found that the security-related features of a website did not correlate with the PCPs' strength. Instead, those websites which were not affected by the consequences of bad usability (e.g. government sites, because users have no alternative), had the strongest PCPs.

However, it has been several years since their investigation and one might wonder: Has the landscape of PCPs on the Internet changed since their initial investigation? Have Internet PCPs become more or less strict? Have the originally-analysed features lost or gained influence on the strength of PCPs? Also, the original study only examined US websites. Thus, it remains an open question whether the features have the same influence on PCP strength of other countries' websites.

We decided to investigate these questions in a replication of Florêncio and Herley's study [5] (original study), extended by the inclusion of a corresponding sample of German websites. Thereby, our goal was not only to revisit the original research questions of Florêncio and Herley [5], but to explore whether a comparison of PCPs over time, and across country borders, yields new findings.

Our results indicate that the US PCPs have become, on average, stronger in the intervening years. For the US sample, most features retain the predictive power with respect to the strength of a website's PCP found in the original study. However, only one of the features used in the original study emerges as a reliable predictor for the German sample: websites facing a potential loss of users due to poor usability are more likely to have weaker PCPs. Furthermore, German websites employ, on average, weaker PCPs than US websites (see Figure 1). In particular, the PCPs used by German banking websites are significantly weaker than those of US banking websites and also exhibit the lowest average PCP strength in the German sample. This finding, combined with the identified predictive factor, "user has choice" may indicate that German banks are especially keen

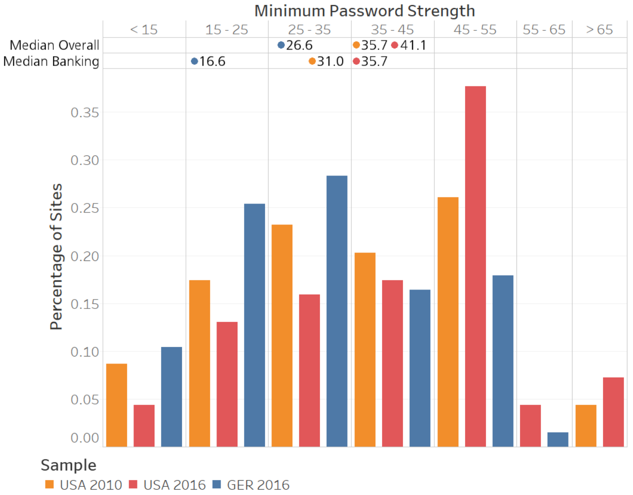

Figure 1: Histogram of PCP strengths (according to the method used in [5]) of the three samples.

Symposium on Usable Privacy and Security (SOUPS) 2017, July 12–14, 2017, Santa Clara, California.

to offer the most user-friendly experience in order to not lose customers due to poor usability.

In the following, we first describe the original study by Florêncio and Herley in detail and present related work. Then, we describe the methodology of this replication study, including our research questions. Thereafter, we present the results for the individual samples in relation to our four research questions. The results are discussed with respect to our research questions afterwards. Finally, we present the conclusions, which we draw from this study.

2. ORIGINAL STUDY

In the original study, Florêncio and Herley [5] investigated the strength of PCPs of 75 different US websites. They sampled websites according to different categories: 15 top traffic websites (determined by Quantcast[1] traffic rank of 20 or higher), 8 high traffic websites (determined by Quantcast traffic rank of 101 to 110), 8 medium traffic sites (determined by Quantcast traffic rank 1001 to 1010), 9 banking websites (the top ranked US banks and brokerages), 10 websites of large universities (determined by their 2006 enrolment numbers), 10 websites of the universities with the top computer science departments (determined as per U.S.News), the 10 government websites with the highest traffic (determined by Quantcast traffic rank), and 7 miscellaneous sites "for comparison interest" [5]. Websites without account systems were simply dropped from their sample. Two websites appeared in two different categories each and were considered in both categories during the analysis.

For all these websites, the PCPs were determined using the following procedure. If possible, an account was created on the website. In case this was not possible, the authors relied on published password policies. To find them, they performed a web search and considered only the first PCP they found.

For each of the identified PCPs, they calculated the minimum strength of the PCP using $N_{min} \cdot \log_2(C_{min})$ where N_{min} is the minimum length allowed by the PCP and C_{min} is the cardinality of the minimum character set required to fulfil the PCP[2]. Their reasoning behind choosing a minimum strength measure is that the intent of PCPs is to enforce a minimum strength among passwords on a website and users usually wont choose passwords which are much more secure. They acknowledge that it is not a perfect measure and does not model guessing resistance, but argue that it preserves the ordering of PCPs in terms of burden on the user. In this paper, we refer to the minimum strength of a PCP without explicitly including the term *minimum* for readability's sake. We simply refer to it as *PCP strength.*

Using this sample of PCPs and their respective strengths, they investigate the effects of the following website features:

1. Observation and evidence with regard to breaches.
2. The size of the service as determined by traffic rank and number of users.
3. Whether the user name is public on the website.
4. The value of the resources protected based on the type of website.
5. The extractable value of the resources protected based on the monetisation of data gained from breaches.
6. Who lives with the consequences of a breach as determined by the policies of the websites.
7. Whether third party advertising is accepted on that website as determined by the Quantcast advertising information.
8. Whether the site advertises as determined by the use of Google Sponsored Links of that website.
9. Whether the user can choose alternative websites offering the same service.

Thereby, they argue that features related to security (i.e. features 1-6) might increase the PCP strength and the features related to attracting users (i.e. features 7-9) might decrease PCP strength.

All their comparisons use the median as measure of central tendency for the strength values. They find that none of the security related features have an effect on PCP strength. However, all features related to attracting users have the anticipated effect, i.e. websites that either advertise themselves, display advertisements of third party websites, or those where users can choose alternatives are more likely to have weaker PCPs.

3. RELATED WORK

Aside from Florêncio and Herley's study, several different aspects of PCPs have been the focus of other studies. In the following, we present selected related research.

Komanduri et al. [9] investigated the security and usability properties of five different constructed (i.e. not taken from the wild) PCPs in an online study. They found that relatively simple password composition policies like mandating at least 16 characters and no other restrictions (i.e. their basic16 policy) yield much better security and usability results than mandating a length of at least 8 characters as well as the usage of uppercase letters, lowercase letters, numbers and symbols (i.e. their comprehensive8 policy).

Follow-up studies using similar constructed PCPs and focusing on the security [8] or both, security and usability, [18, 19] of PCPs could replicate these findings: the typical *comprehensive8* PCPs are among the weakest PCPs in terms of guessing resistance and also exhibit unfavourable usability properties. The authors recommend exchanging any *comprehensive8* PCP in use with one of three alternatives they identify in their study as exhibiting better security and better usability properties.

Focusing on PCPs used by websites in the wild, Kuhn and Garrison [10] conducted a survey of the difference in PCP

[1] www.quantcast.com

[2] Note the following two aspects of the original methodology, which were not explicitly mentioned in [5]: (1) if a password requires special characters, Florêncio and Herley [5] used a cardinality of 34 to calculate the PCP strength, regardless of the number of special characters actually allowed by the PCP, and (2) if a password allowed four character sets, but required only 3, they included letters instead of special characters in all instances.

strength over a period of two years. Their assessment of security was based solely on the minimum length of the password mandated by the PCP and therefore uses a much simpler measure than Florêncio and Herley [5] in their study. They found that more websites used PCPs in 2009 than in 2007 and that the mandatory length of passwords imposed by those PCPs increased in the same time frame.

Employing a similar methodology for the selection of their PCP sample as Florêncio and Herley, Seitz et al. [17] analysed the PCPs of the top 100 German websites (according to Alexa rankings). Their focus however, was to evaluate the potential of password reuse among these sites. They found that despite the great diversity among the PCPs, it is fairly easy to find a password that can be reused on virtually all websites.

Also considering the traffic rank of websites as an important factor, Preibusch and Bonneau [16] find that high traffic rank websites (determined by Alexa ranks) are more likely to attempt to prohibit password sharing among users by blocking listings of credentials at the password sharing community `bugmenot.com`.

Focusing on the practical aspects of password security, Florêncio et al. [6] summarise the password research relevant for system administrators. They present findings relating to the guessing resistance required to withstand offline and online attacks and also discuss implementation details such as appropriate hash functions. They advise to consider in the formulation of PCPs, that offline guessing attacks are much less frequent than originally thought and that online guessing attacks should be the focus when it comes to determining the guessing resistance of passwords.

In another work focused at practitioners, Zhang-Kennedy et al. [25] investigate long standing password management and composition rules. They discuss the viability of these rules based on the results of current research. Based on their findings, they introduce an updated set of password rules, aimed at decreasing the burden on users.

Aiming at increasing the security of existing PCPs in practice, Blocki et al. [2] describe a theoretical model for optimising password composition policies by maximising the PCP's minimum entropy from a set of sample passwords chosen by users.

4. METHODOLOGY

In conducting our replication of the original study by Florêncio and Herley [5], we computed the strength of the PCPs used in 2016 on those US websites used in the original study and re-investigated correlations between the websites' features as identified in [5]. We also applied their approach to a corresponding German sample. We chose a German sample in addition to the US ones for technical reasons: as Germans we can easily understand the PCP descriptions and conduct follow-up studies.

The purpose of our study is to investigate to which extent the results of the original study can be generalised across time and different countries, i.e. for which website features there exist differences between samples and where there exist none. Based on the results of our analyses, we answer the following questions:

RQ1: Has the average PCP strength in the US sample changed since the original study?

RQ2: Do the effects of the website features on the PCP strength from the original study still apply to the USA 2016 sample?

RQ3: How do the German and US samples compare in terms of PCP strength?

RQ4: Do the effects of the website features on the PCP strength from the original study translate to the German sample?

To ensure comparability with the original study, we employed the methodology as used by Florêncio and Herley [5] as closely as possible. However, in order to render our investigation viable and its results meaningful, we needed to adapt the methodology in some respects. In the following, we detail the alterations to the original study's methodology. Where not stated differently, we replicated the original methodology as described in section 2. All PCP data was collected in January 2016.

4.1 Identification of US Website Samples

For the US sample, we used the 75 websites from the original sample of Florêncio and Herley [5]. Five of these were excluded from our investigation due to the following reasons: *highschoolsports.net* was no longer available at the time of the survey, *youtube.com* now uses the google.com user account system, *ask.com* as well as *hollywood.com* seem no longer to have a user account system, and *typepad.com* did not provide information about its PCP on the website. Account creation would have required us to provide payment details. The remaining 70 websites were all included in our investigation. A list of all websites in the US sample can be found in the appendix in Table 5.

4.2 Identification of German Website Sample

We collected a comparative sample of 67 German websites. A list of all websites can be found in the appendix in Table 6. The websites were collected according to the categories defined in the original study: top, high, and medium traffic sites as well as banks, universities and government websites. We did not consider the miscellaneous category for the collection of the German sample due to its ambiguity. In the following, we describe the collection of the German website sample in detail.

German Top, High, and Medium Traffic Websites. Florêncio and Herley used the traffic rank information from the Quantcast service to identify top, high, and medium traffic websites. However, the service does not seem to provide reliable traffic rankings for German websites. While the site offers a list of the top 100 German websites, it does not seem to be representative of actual usage, e.g. none of the search engines included in popular browsers (Google, Bing, and Yahoo) appear. We thus used the alternative Alexa[3] rankings, since their list of German website rankings seemed much more representative and they have also been used in other pertinent studies (e.g. [17, 16]).

[3] `www.alexa.com/topsites/countries/DE`

Analogously to the original study, we used the ranks 1 to 20 and 101 to 110 for the top and high traffic categories respectively. Since the Alexa service only provides the 500 most visited websites, it was impossible to choose the ranks 1001 to 1010 for the medium traffic websites, as chosen in the original study. Therefore, we approximated by using the last ten ranks provided by Alexa[4]. As in the original study, websites without account systems were discarded.

German Banking Websites. For the banking category, we chose several of the largest banks in Germany [4, 11]. Unfortunately, some banks do not offer information about their policies on their websites. Therefore, we could only include four of the ten largest traditional German banks (i.e. banks with brick and mortar branch offices) and five of the ten largest German online-only banks (i.e. banks with no branch offices and only an online presence).

German University Websites. The German university websites used in this study represent the largest German universities, based on official government statistics [22]. The websites of the best-rated computer science departments are based on the CHE university ranking [24].

German Government Websites. Florêncio and Herley used the ten highest traffic websites with a `.gov` top-level domain. Germany does not have an equivalent to the US `.gov` domain, so we resorted to a different identification procedure. First, we manually identified the government websites on the Alexa list of the 500 most frequently-visited websites. This yielded only 5 government websites with an account system. To gather additional government websites for our sample we consulted a report comprising an extensive list of German government websites [14]. Using this list, we were able to identify an additional three websites with user account systems.

4.3 Identification of PCPs

The approach we took in order to identify the PCPs of both samples (USA 2016 and GER 2016) was similar to the one applied in the original study. Where possible, we created an account on the website[5]. Sometimes we used demo accounts to check the password policies. If neither of these approaches was possible, a web search was used to locate the PCP. For the US websites with multiple account systems, we used the PCPs for the same account systems as used in the original study, to support a meaningful comparison. On German websites with multiple account systems, we used the first PCP we found (analogously to the original study).

4.4 Website Accepts Advertising

We collected data regarding the placement of advertisements on the website. The original study relied on the advertising info provided by the Quantcast service to decide whether websites displayed advertisements from third party websites. This information is no longer available. We applied a manual approach. We visited each of the websites and navigated through the pages. If we found any advertisements, we categorised the respective website as having this feature, otherwise not.

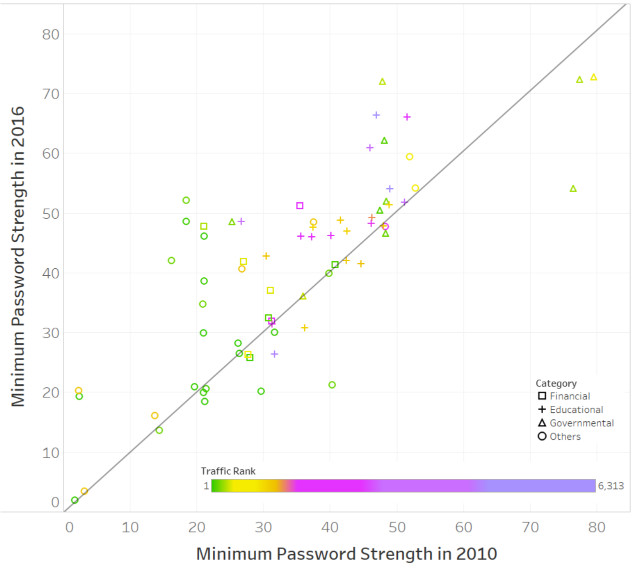

Figure 2: Scatterplot showing the difference in PCP strength (as determined by minimum password strength) for the US sample over time. Each point represents one website. The websites above the diagonal have adopted a stricter PCP since 2010, the websites below the diagonal have adopted a more lenient PCP.

5. RESULTS

In this section, we first present the results for the PCP's strength and then the results for the nine features[6] from the original study and their effects. Note that we refer repeatedly to *the three samples*, which relates to our German website sample (GER 2016) and our updated US website sample (USA 2016) as well as the US website sample from the original study (USA 2010). All correlations reported here were calculated using Pearson's correlation coefficient r using the R statistics system. Table 1 gives an overview of the investigated effects, the results of the original study, and the results of our own investigation.

5.1 Strengths of PCPs

The average strength of the US sample has grown significantly from 35.7 bits in 2010 to 41.4 bits in 2016, whereas the maximum values have declined from 79.0 bits to 71.5 bits respectively. The increase in average PCP strength is caused by 37 websites adopting a stronger PCP than in 2010, while only 8 websites adopted a weaker PCP. The remaining 25 websites did not change their PCP from 2010 to 2016. Figure 2 depicts the change in PCP strength over time in the US samples.

[4] Note, we can report Alexa ranks for the websites of the other categories in our sample, since it is possible to query the Alexa database for any website to get its rank. The identification of the medium traffic websites is hindered since the inverse is not possible (i.e. querying the Alexa database with a specific rank to get the corresponding website).

[5] Note, in some cases, it was not possible to create an actual account, but instead only to carry out the registration process. While no account was created, this process still gave us access to the PCP.

[6] We wont motivate the individual features in this work. Instead, we would like to refer the interested reader to the description in the original study [5] for a detailed description explaining the selection.

Table 1: Overview of the investigated website features and their hypothesised as well as actual effects on PCP strength. "↑" indicates an increase in strength, "↓" indicates a decrease in strength, "-" indicates no effect.

	Hypothesised effect	Actual effect on PCP strength		
Website feature	on PCP strength [5]	USA 2010	USA 2016	GER 2016
Observation and evidence	↑	-	-	-
Size of the service		-	-	-
User name public		-	-	-
Value of the resources protected		-	-	-
Extractable value of the resources protected		-	-	-
Who lives with the consequences of a breach		-	-	-
Advertising accepted	↓	↓	↓	-
Site advertises		↓	-	-
User has choice		↓	↓	↓

Table 2: The median PCP strengths of the websites in the three samples. German websites generally employ weaker PCPs. In particular German banking websites stand out, exhibiting the lowest average in all three samples.

		Traffic			Website type			
Sample	Overall	Top	High	Medium	Bank	University	Government	Others
USA 2010	35.7	19.9	19.9	36.2	31.0	41.7	47.6	19.9
USA 2016	41.4	26.6	41.5	46.5	35.7	47.6	52.7	29.9
GER 2016	26.6	26.6	25.8	19.9	16.6	30.8	47.6	26.6

In comparison to the USA 2016 sample, the German 2016 sample shows a much smaller median of 26.6 bits and also a slightly lower maximum of 59.3 bits. The minimum is equal for all samples (3.3 bits), due to Wikipedia being present in all samples. Figure 1 depicts the distribution of the PCP strengths for all three samples. While the range decreased in the US samples from 2010 (75.7 bits) to 2016 (68.1 bits) and is even smaller for the German sample (55.9 bits), it remains very large in all samples.

Table 2 shows the differences between the samples for the different categories of websites distinguished in the original study. It becomes apparent, that the average PCP strengths have increased in the US sample for all categories. When comparing the the two samples from 2016, German websites employ on average weaker PCPs in every category. Figure 6 in the appendix illustrates the distributions of the PCP strength per category of all three samples.

5.2 Observation and Evidence

One might argue that website providers learn from past events and derive the PCPs from past experiences. Florêncio and Herley approached this question in the original study based on argumentation and while their arguments still hold today, we also applied an evidence-based approach in our investigation using information related to whether the US websites had been affected by a password-related breach or leak in the years between 2010 and 2016. To that end, we conducted web searches for each website in the US sample, in order to identify whether it had been affected by a password breach or leak since the original study. We used the Google search engine with the search terms "password breach", "password leak", "password hack", and "password incident", each in combination with the respective website's name. If we found a security incident exposing password data[7] on the first five pages of search results, we classified a website as having been victim of a breach or leak. Table 5 in the appendix shows the individual classification of each website in the US sample. This classification is, admittedly, only an approximation. Not all leaks are made public, which decreases the precision of our approach.

Using the classification, we split the websites into three categories: those having increased the the strength of their PCPs in the time between 2010 and 2016, those having reduced the strength of their PCPs in that time frame, and those with no change in their PCPs. We hypothesised that if websites operated on their past evidence, then websites that had been the target of a breach since 2010 would be more likely to have increased the strength of their PCPs. Table 3 shows the frequencies of websites classified as detailed above.

We conducted a Fisher's exact test to investigate the effect of websites being affected by a breach on the PCP strength of those websites. It yielded no significant results (FET: $p = 0.415$). Thus, the hypothesis that past breaches have an effect on PCP strength has to be rejected.

5.3 Size of the Service

Florêncio and Herley [5] hypothesised that PCP strength correlates with (a) the size of a website (as determined by the number of user accounts on that site) and (b) the traffic generated by the website (as determined by the Quantcast traffic rank, see section 2 for details). They reject these hypotheses based on the observation that top-traffic services

[7] Indirect attacks such as abusing reset mechanisms were not considered a breach or leak in the sense of our investigation since the actual password is not revealed when security questions are easily guessed. Also, attacks leaking passwords in the clear (such as phishing) were not considered, since stronger passwords do not protect against these kinds of attacks.

Table 3: Frequencies of websites classified along the two characteristics: (a) whether a website has been victim of a breach or not and (b) whether the website uses a stronger, weaker, or unchanged PCP.

	PCP is		
	Stronger	Unchanged	Weaker
Breach	13	6	1
No breach	24	19	7

with many users (such as Facebook or GMail) have much weaker PCPs than universities which have significantly lower traffic ranks and also lower numbers of users (approximated from undergraduate enrolment). Since the traffic ranks in 2016 do not match the original sampling and in some cases deviate significantly from their 2010 ranks (e.g. Myspace had rank 16 in 2010 and ~1000 in 2016), a direct comparison is not possible.

However, since the original study, the top traffic websites have increased their number of users, e.g. Facebook from ~400 million to ~1700 million [21] or GMail from 91 million to 1000 million [20]. In contrast, in the same time frame the number of student enrolments remained steady for the lower traffic examples used by Florêncio and Herley [5], e.g. Ohio State University 51800 in 2010 and 51759 in 2016 [15]. Hence Florêncio and Herley's argument seems to hold. However, it must be acknowledged that the measure chosen by Florêncio and Herley for the approximation of the number of university user accounts (i.e. undergraduate student enrolments) might not be optimal (see section 7 for a discussion of this limitation).

In addition, we conducted a correlation analysis for our German sample based on the Alexa ranks. We found a weak negative correlation between the Alexa ranks of the websites in the German sample and the strength of their PCPs ($r = -0.16$). Thus, our results support the findings of the original study for the German sample as well.

5.4 User Name Public

When user names are publicly available, bulk guessing attacks, where attackers try only the most frequent passwords for all accounts known to them, become much more viable. Therefore, Florêncio and Herley hypothesised that websites with public user names might employ PCPs with a higher average strength. They assumed social networks', auction websites' and email providers' user names to be public and stated that for universities the user name is often public as well. For the US sample, the findings from the original study can be directly transferred: a Wilcoxon rank sum test results in rejecting the hypothesis that there is a difference in PCP strength between websites with public user names and websites where user names are not publicly accessible ($W = 648.5, p = 0.674$).

For the German sample it is of note that university user names do not seem to be public in general. Some universities from our sample use the student id number (e.g. FernUniversität Hagen), the student email address (e.g. Goethe Universität Frankfurt) or personal information such as first and last name (e.g. Universität Köln), and this might make bulk guessing easier. However, there are also universities in our sample which explicitly use random user names (e.g. TU Darmstadt or Universität des Saarlandes). In particular, for most German universities this information is not publicly available. However, even with this difference, a Wilcoxon rank sum test results in a rejection of the hypothesis that there is a difference in PCP strength between websites with public user names and websites where user names are not publicly accessible ($W = 456.5, p = 1.000$). Our findings fully support the findings of the original study for all samples.

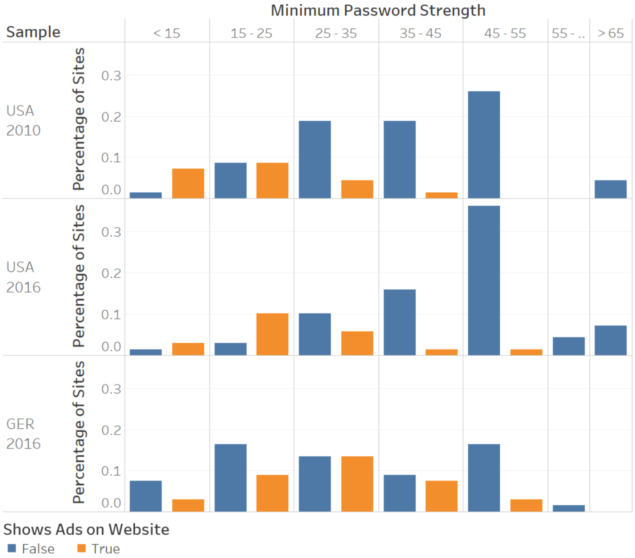

Figure 3: Histograms for the three samples, showing the distributions for websites that display third party advertisements and those that do not.

5.5 Value of the Resources Protected

While investigating the values of the resources protected in the USA 2016 sample, the same trend already described in the original study became apparent. Financial services have, on average, more lenient PCPs than government websites. In particular, the same example from the original study still holds. Both, Fidelity (increase in PCP strength since 2010) and Paypal (no difference in PCP strength since 2010), still have weaker PCPs than USAJobs (no difference since 2010). Table 2 shows the average PCP strengths of banking websites for all samples. German banking websites have the lowest average PCP strengths of all three samples (16.6 bits). US banking websites have significantly higher average PCP strengths in both 2010 and 2016.

5.6 Extractable Value of the Resources Protected

Florêncio and Herley hypothesise that the extractable value of user accounts might increase the PCP strength of respective websites. To identify the websites considered the most valuable, they consider those heavily targeted by phishers, since they argue that these represent the websites whose accounts offer the best monetisation. In contrast to their hypothesis, they find that the most phished brands in 2009 all have relatively low strength PCPs.

For the extractable value of the resources protected by the respective passwords, we see the same effects as described

Table 4: The median PCP strengths of the websites in the three samples in relation to whether they accept third party advertisements, whether the websites advertise themselves, and whether the user can choose alternative websites.

	Accepts ads		Advertises		User choice	
Sample	Yes	No	Yes	No	Yes	No
USA 2010	19.9	41.1	31.0	35.7	19.9	41.6
USA 2016	19.9	47.6	47.6	41.4	26.6	47.6
GER 2016	26.6	26.6	22.9	26.6	26.2	31.0

in the original study. According to the APWG [1], financial websites are still among the ones most heavily targeted by phishers. Close to 19% of phishing attacks target this sector. In contrast, less than 2% of attacks target government and education websites, both of which have much higher average PCP strengths.

5.7 Who Lives with the Consequences of a Breach

When a service has to compensate users for possible consequences, this financial threat could be a reason for website providers to enforce stronger PCPs. As noted in the original study, this was not the case in 2010. Our investigation provides even more evidence in this regard. Banks are still among the websites employing weak PCPs in all samples (in particular in the German sample). Yet they often compensate users for unauthorised transactions [5]. In Germany, account holders only have to cover the first 150€ themselves. Our investigation fully supports the original study's findings.

5.8 Advertising Accepted

Some websites generate their revenue through third party advertisements. Table 4 shows the median PCP strengths for the websites displaying third-party advertisements and the ones that do not. The US samples underline the findings of the original study: a Wilcoxon rank sum test indicated that websites displaying third-party advertisements had significantly weaker PCPs than those that did not display advertisements in the USA 2016 sample ($W = 759.0, p < 0.001$). It is interesting to note that the median PCP strengths for the websites accepting advertising did not change from 2010 to 2016. The overall increase in the USA 2016 sample stems solely from websites not displaying advertisements.

However, for the German sample, displaying adverts does not seem to have a significant effect on the average PCP strength. The median values for both groups of websites are the same as the overall median strength of 26.6 bits already reported in section 5.1. Consequently, a Wilcoxon rank sum test results in a rejection of the hypothesis that there is a difference in PCP strength between websites displaying third-party adverts and those that do not display third-party adverts in the German sample ($W = 554.5, p = 0.617$). Figure 3 illustrates this effect across all three samples.

5.9 Site Advertises

To generate traffic, some websites place advertisements on other websites. As an indicator of whether websites place such ads, we use (analogously to the original study) Google sponsored links. Figure 4 shows the distributions of PCP strength of websites utilising Google sponsored links and those that do not do this for all three samples. Again, there is no visible effect for this feature in the German sample; both median PCP strength values are almost identical at 26.6 bits for the non-advertising sites and 26.3 for the advertising sites (cf. Table 4). A Wilcoxon rank sum test supports this finding, resulting in a rejection of the hypothesis that there is a difference between advertising and non-advertising websites in the German sample ($W = 505.0, p = 0.366$).

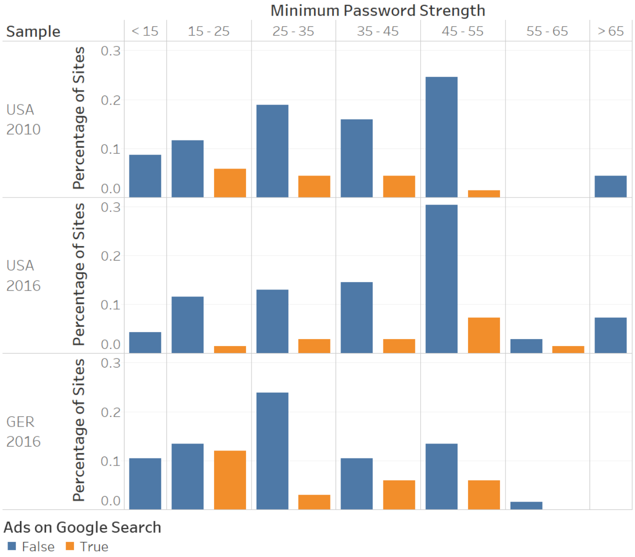

Figure 4: Histograms for the three samples, showing the distributions for websites that are advertising using sponsored links on Google Search and those that do not advertise using sponsored links.

In the USA 2016 sample, a Wilcoxon rank sum test results in rejecting the hypothesis that there is a difference in PCP strength between advertising and non-advertising websites as well ($W = 297.0, p = 0.667$). However, while not significant, the results might indicate a weak reversal effect, illustrated by Figure 4: in 2010, non-advertising websites had the higher average PCP strength, in 2016 the advertising websites have the higher average PCP strength (cf. Table 4).

5.10 User Has Choice

Concerning whether the user has a choice to use the website, the results are consistent for all three samples. Wilcoxon rank sum tests indicated that websites where users can choose between alternatives have significantly weaker PCPs than those where users have no choice for both, the USA 2016 sample ($W = 976.5, p < 0.001$) as well as the German sample ($W = 780.0, p = 0.004$). When comparing the two samples from 2016, German websites without alternatives for the user, unsurprisingly, have a lower PCP strength than the corresponding US websites. On the other hand, there is no difference between the websites of both samples where users can choose alternatives. Thus, the feature seems to have a similar effect on the strength of the PCPs in all three samples.

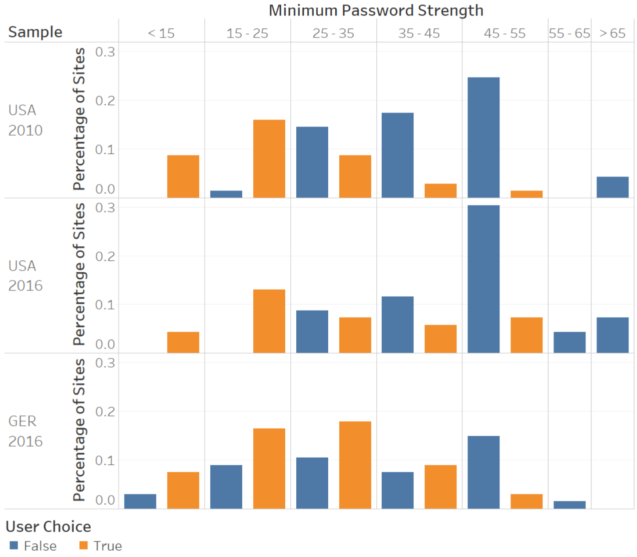

Figure 5: Histograms for the three samples, showing the distributions for websites where users can choose an alternative website and where users have no choice.

6. DISCUSSION

In our replication study, we re-investigated the effects of several website features on the strength of websites' PCPs. The goal of this replication was not only to revisit the original research questions of Florêncio and Herley [5], but to explore whether a comparison of PCPs over time and across country borders yields new findings. We discuss the findings related to each of our research questions in the following.

6.1 Has the average PCP strength in the US sample changed since the original study?

Based on the data from our replication of Florêncio and Herley's study [5], the answer to the first research question is a definitive "yes". The average PCP strength in the US sample has risen from 35.7 bits to 41.4 bits since the original study. 52.9% of the websites use 2016 stronger PCPs than they did in 2010. In contrast, only 11.4% of the websites used weaker PCPs.

While this trend supports similar findings by Kuhn and Garrison [10], it contradicts established expert opinion: NIST's newly drafted rules regarding password security [7] recommend using a PCP with *"at least 8 characters"* and *"no other complexity requirements for memorized secrets"*. Such a PCP has a minimum strength of 26.56 bits. The US average from 2016 is at 41.1 bits 35.8% higher than this recommendation. Therefore, it seems that PCPs found in the wild are much more complex (as determined by Florêncio and Herley's measure) than what is recommended. Such overly complex PCPs might be, usability-wise, alarming.

The reasons for this rise in PCP strength, however, cannot be identified from the data collected in our study. As in the original study, all hypotheses regarding factors increasing PCP strength had to be rejected. Thus, no explanation for the rise in PCP strength emerges from the original hypotheses. Furthermore, our additional investigation into the effects of breaches on PCP strength reveals that the observable rise in PCP strength between 2010 and 2016 cannot be attributed to the website being breached either. Consequentially, other features must be the driving force behind the rise in PCP strength. While it might be that website providers try to counter increasing attacker capabilities for offline attacks by employing stronger PCPs, this would contradict Florêncio et al's . [6] recommendations to focus on online guessing when designing PCPs, which is better addressed with lock-out policies. Yet the identification of the influential features, in this regard, constitutes one important focus for future work.

6.2 Do the effects of the website features on the PCP strength from the original study still apply to the USA 2016 sample?

Regarding the answer to our second research question, the results from our study concur with the findings of the original study for all features except one. All website features, which Florêncio and Herley hypothesised to increase PCP strength, still do not have that effect. Also, websites that display third party advertisements and websites where users can choose alternatives still have significantly weaker PCPs.

The only divergence from the findings in the original study is related to whether websites advertise to attract users. This feature seems to have lost its effect on the PCP strength. The reasons for this divergence, however, remain unclear.

6.3 How do the German and US samples compare in terms of PCP strength?

With respect to the third research question, the answer we can give from the results of our study is that websites in the German 2016 sample employ in every category on average weaker PCPs than those in the USA 2016 sample. For the three categories medium-traffic, banking, and education, the websites in the German 2016 sample have even lower average PCP strengths than the websites of those categories in the US sample had in 2010. Especially German banking websites stand out in this regard: While the passwords on these websites protect the most (monetary) value, they are created under the PCPs exhibiting the lowest average strength across all three samples.

However, one important aspect regarding German banking websites is that they implement two-factor transaction authorisation. The notable difference to two-factor authentication is that users can log in (authenticate) without the second factor, but authorising transactions requires a second factor (usually a so-called TAN, a transaction number for one-time use delivered either in advance e.g. as a list of TANs via mail or nowadays on demand e.g. via smartphone apps). Thus, the actual extraction of resources requires more than mere knowledge of the password, but carrying out this kind of attack is not impossible [12]. To gain further insight, we contacted a local bank. Their perspective is that tight lock-out policies and high security data centres made strong PCPs unnecessary. Hence, the trade-off of employing a lower-strength PCP in conjunction with tight lock-out policies and two-factor authorisation might be, usability-wise, a favourable trade-off. Whether users agree with this perspective and find such lower strength PCPs adequate for protection in the banking context remains an open question that we cannot answer without further investigation.

6.4 Do the effects of the website features on the PCP strength from the original study translate to the German sample?

The results of our study indicate that the effect of only one feature translates to the German sample. User choice is the only feature affecting the PCP strength in the German sample. When a user can choose alternatives to a certain website, that website is more likely to employ a weaker PCP. However, in contrast to the US samples, the display of adverts does not seem to have a significant effect on PCP strength in the German sample. Neither the display of advertisements, nor using adverts to attract users, has a significant effect on the PCP strength of websites in the German sample. Our data does not suggest any explanations for this. Regarding the effects of features hypothesised to increase PCP strength, all hypotheses had to be rejected. This confirms the findings of the original study.

Therefore, we argue that only two factors truly influence the strength of a website's PCP across the three samples: (1) a general tendency to enforce PCPs which are as strong as possible, and (2) the dependence on usability to attract users, leading to weaker policies. As already pointed out by Florêncio and Herley in the original study, and further supported by the findings of this replication, this trade-off is decided by websites more or less off-the-cuff. This holds for the US and German samples. To illustrate for the US sample: the range among US universities is 39.2 bits in 2016. However, it is unclear, why Princeton (PCP strength of 65.8 bits) should feel the need to enforce a significantly stronger PCP than Northwestern University (PCP strength of 26.6 bits). With respect to the German sample, the average strength of the PCPs seems, with 26.6 bits, to be very close to NIST's recommendation (26.56 bits). However, the large range of 55.9 bits across PCP strengths and, in particular, high strength PCPs on seemingly low value sites (e.g. 47.6 bits for the news site `spiegel.de`) give rise to doubt regarding a more systematic approach being applied to PCP choice on German websites.

7. LIMITATIONS

As already acknowledged by Florêncio and Herley [5], the minimum PCP strength measure employed in their study can only serve as a rough estimation and more precise measures of guessing resistance exist. However, as Florêncio and Herley pointed out in the original study, their measure is not intended to model resistance to guessing attacks, but only complexity of the resulting passwords. Since we adopt this measure to perform our replication, this limitation applies to our study as well.

We also decided to not use any additional measures of password security, since any reliable estimate of a PCPs strength (e.g. α-guesswork [3] or guess numbers [23]) would require collecting passwords created under the respective PCP. However, collecting adequate numbers of passwords for the calculation of these measures is beyond the scope of this work.

Another limitation that arises from adopting the original methodology and the nature of performing a replication is that PCP strength is only investigated in relation to the website features. The effects of other influencing factors such as technologies employed by the user (e.g. two-factor authentication, password managers, etc.) are not considered. Especially, two-factor authentication might play a role in some categories: as explained before in section 6.3 German banks require the user to provide a second factor to authorise transactions. It must be assumed that this influences the PCP choice of banks. However, other factors might play a role here as well. For example, traditional banks (i.e. banks with brick and mortar branch offices) might have relatively strict lockout policies, since their customers can simply visit the local branch office to get their account unlocked. Therefore, while future study designs should include these interesting extensions of the methodology and consider such technologies, it was beyond the scope of this replication study.

The third limitation of our replication study is also shared with the original study. For the identification of the PCPs in our samples, we followed the same methodology as Florêncio and Herley. Thus, we also created an account at the website whenever possible (the information whether an account was created is available for each of the websites in our samples in tables 5 and 6 in the appendix). However, when this was not possible we also followed the methodology of the original study and conducted a web search. This can lead to imprecisions in the samples, since sometimes the found PCPs might represent guidelines not enforced during the actual password choice or for a university might only be enforced for a specific account system, but not for others.

The final limitation that our replication study shares with the original study is the approximation of user accounts at universities by undergraduate enrolment numbers. Using purely these numbers might not be optimal, since the number of accounts managed by universities nowadays might only loosely correlate with the number enrolled undergraduate students due to the emergence of other account systems at the universities (e.g. affiliated research institutes, alumni, donors, or even accounts for the purchase of sports tickets)[8]. However, we argue that this does not affect our results, since even if the undergraduate enrolment is not fully representative of the number of user accounts at universities, it is unlikely that universities reach the numbers of users of the top traffic websites.

A limitation arising from the longitudinal analysis is that we decided to use the same websites as the original study for the USA 2016 sample instead of collecting a new sample from the same categories. Therefore, some of the websites now belong to a different category. However, we decided to use the same website, since this affects less than 1/10 of the sample and we believe the longitudinal comparison (enabled only by using the same websites) adds special value to this paper.

Lastly, the additional breach analysis we conducted (cf. section 5.2) should only be treated as an approximation. While some countries have passed laws mandating the reporting of data breaches (cf. e.g. [13]), this does not hold for all jurisdictions and consequentially not all leaks are made public. Moreover, our search terms might have been insufficient to identify all available information on breaches at the respective sites.

8. CONCLUSION

In this paper, we presented a replication of the study by Florêncio and Herley [5]. Thereby, the contribution of our

[8] Thanks to reviewer 1 for pointing this out.

paper is twofold: (1) the comparison of password composition policies in the US sample over time, and (2) the comparison of password composition policies across country borders (i.e. between Germany and the US).

Regarding the first contribution, it became apparent that US PCPs have become, on average, stronger and that all but one website feature have retained their effects on PCP strength in the intervening years. While the former is in line with the findings of similar studies [10], it contradicts established expert opinion and might be, in terms of usability, an alarming finding. Moreover, our results indicate that two website features correlate with decreased PCP strength in the USA 2016 sample (i.e. "advertising accepted" and "user has choice"), but none of the website features seem to correlate with increased PCP strength in practice. Therefore, future work is needed to identify the reasons behind the rise in PCP strength in the US from 2010 to 2016. With respect to the effects of the website features on PCP strength in the US samples, only the effect associated to whether websites advertise to attract users seems to have changed. The effect could not be found in the 2016 sample.

Regarding the second contribution, we observed, on average, lower PCP strengths in the German sample than in the US samples. German banks stand out as having particularly weak PCPs. Together with the fact that "User has choice" emerges as the only website feature exhibiting an effect in the German sample, it seems that German banks are especially keen to maximise usability and optimise the user experience. They provide the user with a favourable trade-off by combining tight lock-out policies with the requirement of a second factor to authorise transactions. However, whether users consider this trade-off adequate for the banking context, or might even want to make similar trade-offs in other contexts as well, is open for future investigation.

9. ACKNOWLEDGMENT

We would like to thank Karen Renaud, Benjamin Reinheimer and Annika Hilt for their valuable input and feedback. Furthermore, we thank the anonymous reviewers and our shepherd Mary Ellen Zurko for their construcutve and helpful feedback.

This work has been developed within the project 'KMU AWARE' which is funded by the German Federal Ministry for Economic Affairs and Energy under grant no. BMWi-VIA5-090168623-01-1/2015. Furthermore, it has been supported in part by the German Federal Ministry of Education and Research (BMBF) within CRISP (www.crisp-da.de/). The authors assume responsibility for the content.

10. REFERENCES

[1] Anti-Phishing Working Group. Phishing Activity Trends Report - 1st Quarter 2016. Technical report, 2016.

[2] J. Blocki, S. Komanduri, A. Procaccia, and O. Sheffet. Optimizing password composition policies. In *EC '13: Proceedings of the fourteenth ACM conference on Electronic commerce.* ACM Request Permissions, June 2013.

[3] J. Bonneau. The Science of Guessing: Analyzing an Anonymized Corpus of 70 Million Passwords. In *IEEE Symposium on Security and Privacy*, pages 538–552, 2012.

[4] Bundesverband deutscher Banken e.V. Zahlen, Daten, Fakten der Kreditwirtschaft. `https://bankenverband.de/publikationen/zahlen-daten-fakten/`. Accessed: 2016-01-16.

[5] D. Florêncio and C. Herley. Where do security policies come from? In *SOUPS '10: Proceedings of the Sixth Symposium on Usable Privacy and Security*, page 1. ACM Press, 2010.

[6] D. Florêncio, C. Herley, and P. C. van Oorschot. An Administrator's Guide to Internet Password Research. In *Large Installation System Administration Conference*, pages 35–52, 2014.

[7] P. A. Grassi, J. L. Fenton, E. M. Newton, R. A. Perlner, A. R. Regenscheid, W. E. Burr, J. P. Richer, N. B. Lefkovitz, J. M. Danker, Y.-Y. Choong, K. K. Greene, and M. F. Theofanos. NIST Draft SP 800-63B: Digital Identity Guidelines - Authentication and Lifecycle Management. `https://pound.netzpolitik.org/wp-upload/Analyse-staatlicher-Websites-Bewertung.pdf`. Accessed: 2016-01-05.

[8] P. G. Kelley, S. Komanduri, M. L. Mazurek, R. Shay, T. Vidas, L. Bauer, C. Wiedeman, L. F. Cranor, and J. Lopez. Guess again (and again and again): Measuring password strength by simulating password-cracking algorithms. In *2012 IEEE Symposium on Security and Privacy*, pages 523–537. IEEE, 2012.

[9] S. Komanduri, R. Shay, P. G. Kelley, M. L. Mazurek, L. Bauer, C. Wiedeman, L. F. Cranor, and S. Egelman. Of Passwords and People: Measuring the Effect of Password-Composition Policies. In *CHI '11: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 2595–2604, New York, New York, USA, 2011. ACM Press.

[10] B. T. Kuhn and C. Garrison. A survey of passwords from 2007 to 2009. In *Information Security Curriculum Development Conference*, pages 91–94, New York, New York, USA, Sept. 2009. ACM.

[11] Modern Banking. Die größten Direktbanken gemessen an der Kundenzah. `http://www.modern-banking.de/marktanteil_direktbanken.htm`. Accessed: 2016-01-16.

[12] C. Mulliner, R. Borgaonkar, P. Stewin, and J.-P. Seifert. SMS-Based One-Time Passwords: Attacks and Defense. In *Detection of Intrusions and Malware, and Vulnerability Assessment*, pages 150–159. Springer, 2013.

[13] National Conference of State Legislatures. Security Breach Notification Laws. `http://www.ncsl.org/research/telecommunications-and-information-technology/security-breach-notification-laws.aspx`. Accessed: 2017-02-22.

[14] Netzpolitik.org. Trackingtools auf Websites staatlicher Institutionen. `https://pound.netzpolitik.org/wp-upload/Analyse-staatlicher-Websites-Bewertung.pdf`. Accessed: 2016-01-20.

[15] Ohio State University Institutional Research and Planning. Statistical Summary. `https://www.osu.edu/osutoday/StatisticalSummary2015.pdf`, 2015.

[16] S. Preibusch and J. Bonneau. The Password Game:

Negative Externalities from Weak Password Practices. In *International Conference on Decision and Game Theory for Security*, pages 192–207, Berlin, Heidelberg, Nov. 2010. Springer, Berlin, Heidelberg.

[17] T. Seitz, M. Hartmann, J. Pfab, and S. Souque. Do Differences in Password Policies Prevent Password Reuse? In *CHI Conference Extended Abstracts*, pages 2056–2063, New York, New York, USA, May 2017. ACM.

[18] R. Shay, L. F. Cranor, S. Komanduri, A. L. Durity, P. S. Huh, M. L. Mazurek, S. M. Segreti, B. Ur, L. Bauer, and N. Christin. Can long passwords be secure and usable? *Proceedings of the 32nd annual ACM conference on Human factors in computing systems - CHI '14*, 2014.

[19] R. Shay, S. Komanduri, A. L. Durity, P. S. Huh, M. L. Mazurek, S. M. Segreti, B. Ur, L. Bauer, N. Christin, and L. F. Cranor. Designing Password Policies for Strength and Usability. *ACM Transactions on Information and System Security (TISSEC)*, 18(4):13–34, 2016.

[20] statista. Number of active Gmail users worldwide from January 2012 to February 2016 (in millions). `https://www.statista.com/statistics/432390/active-gmail-users/`. Accessed: 2016-02-29.

[21] statista. Number of monthly active Facebook users worldwide (in millions). `https://www.statista.com/statistics/264810/number-of-monthly-active-facebook-users-worldwide/`. Accessed: 2016-02-29.

[22] Statistisches Bundesamt. Hochschulen. `https://www.destatis.de/DE/ZahlenFakten/GesellschaftStaat/BildungForschungKultur/Hochschulen/Hochschulen.html`. Accessed: 2016-01-17.

[23] B. Ur, S. M. Segreti, L. Bauer, N. Christin, L. F. Cranor, S. Komanduri, D. Kurilova, M. L. Mazurek, W. Melicher, and R. Shay. Measuring real-world accuracies and biases in modeling password guessability. In *USENIX Security Symposium*, 2015.

[24] Zeit Campus. CHE Hochschulranking 2015/16. `http://ranking.zeit.de/che2015/en/`. Accessed: 2016-01-17, now replaced with the 2016/17 version: `http://ranking.zeit.de/che2016/en/`.

[25] L. Zhang-Kennedy, S. Chiasson, and P. van Oorschot. Revisiting Password Rules: Facilitating Human Management of Passwords. In *Information Assurance and Security Workshop*, 2016.

Table 5: The US website sample (USA 2016) comprising 70 websites. Traffic ranks according to Quantcast.

Website	Traffic	Account	Min.	Size	Min. Strength		Accepts	Places	User	Affected
	Rank	Created?.	Length	Charset	2016	2010	Ads?	Ads?	Choice	by Breach
Top Traffic Sites										
Google	1	y	8	10	26,6	26,6	y	n	y	n
Facebook	2	y	6	10	19,9	19,9	y	n	y	n
Yahoo	3	y	9	10	29,9	19,9	y	n	y	y
AOL	6	y	8	10	26,6	26,6	y	n	y	y
Live	8	y	8	36	41,4	19,9	y	n	y	y
Wikipedia	9	y	1	10	3,3	3,3	n	n	y	n
eBay	10	y	6	36	31,0	31,0	y	n	y	y
Amazon	11	y	6	10	19,9	19,9	y	y	y	y
weather	13	y	6	10	19,9	19,9	y	n	y	n
answers	15	y	6	10	19,9	3,3	y	n	y	n
Myspace	16	y	6	10	19,9	31,0	n	n	y	n
Craigslist	17	y	8	26	37,6	19,9	n	n	y	n
adobe	20	y	8	62	47,6	19,9	n	y	y	y
High Traffic Sites										
nih.gov	101	n	8	62	47,6	47,6	n	n	n	n
capitalone.com	102	n	8	36	41,4	41,4	n	y	n	n
rockyou.com	103	y	6	10	19,9	41,4	y	n	y	n
overstock.com	107	y	8	36	41,4	16,6	n	n	y	n
latimes.com	108	y	7	36	36,2	19,9	y	n	y	n
intuit.com	109	y	8	96	52,7	19,9	n	y	y	y
cbssports.com	110	y	4	10	13,3	13,3	y	n	y	n
Medium Traffic Sites										
wowwiki.com	1001	y	1	10	3,3	3,3	y	n	y	n
virginia.edu	1002	n	8	62	47,6	36,2	n	n	n	n
pgatour.com	1003	y	6	10	19,9	3,3	y	n	y	n
mit.edu	1006	n	8	36	41,4	31,0	n	n	n	n
okcupid.com	1007	y	5	10	16,6	13,3	y	n	y	n
istockphoto.com	1008	y	8	36	41,4	25,8	n	y	y	n
Banks										
Fidelity	224	n	8	62	47,6	19,9	n	y	n	n
Vanguard	629	n	8	10	26,6	26,6	n	n	n	n
Schwab	2266	n	6	36	31,0	31,0	n	y	n	n
WellsFargo	80	n	6	36	31,0	31,0	n	n	n	n
BoA	48	n	8	36	41,4	41,4	n	n	n	n
J P Morgan Chase	2186	n	8	86	51,4	36,2	n	n	n	n
Citibank	316	n	6	62	35,7	31,0	n	n	n	n
PayPal	29	y	8	10	26,6	26,6	n	n	y	n
US Bank	316	n	8	36	41,4	26,6	n	n	n	n
Large Universities										
Ohio State U	1811	n	8	62	47,6	41,4	n	y	n	n
Arizona State U	3288	n	10	62	59,5	47,6	n	y	n	y
U. of Florida	1382	n	8	62	47,6	47,6	n	n	n	n
U. of Minn.	919	n	6	36	31,0	35,7	n	n	n	n
U. of Texas	946	n	8	62	47,6	47,6	n	n	n	n
U. of Central Florida	6313	n	8	96	52,7	47,6	n	n	n	y
Michigan State U.	1174	n	8	62	47,6	47,6	n	n	n	n
Texas A & M	1418	n	8	62	47,6	35,7	n	n	n	y
U. South Florida	2364	n	8	62	47,6	35,7	n	n	n	n
Penn. State U.	977	n	8	36	41,4	41,4	n	n	n	y
Universities with top CS departments										
MIT	1006	n	8	36	41,4	31,0	n	n	n	n
Stanford	858	n	8	96	52,7	47,6	n	n	n	y
UC Berkeley	905	n	9	36	46,5	41,4	n	n	n	n
CMU	3651	n	8	96	52,7	52,0	n	n	n	n

Continued on next page

Table 5 – continued from previous page

Website	Traffic	Account	Min.	Size	Min. Strength		Accepts	Places	User	Affected
	Rank	Created?.	Length	Charset	2016	2010	Ads?	Ads?	Choice	by Breach
UIUC	3384	n	8	62	47,6	26,1	n	n	n	n
Cornell	955	n	8	62	47,6	41,7	n	n	n	y
Princeton	1879	n	10	96	65,8	52,7	n	n	n	y
U. of Washington	1032	n	8	36	41,4	45,6	n	n	n	n
Georgia Tech.	4687	n	11	62	65,5	47,6	n	n	n	n
U. of Texas	946	n	8	62	47,6	47,6	n	n	n	y
Government Sites										
irs.gov	63	n	8	70	49,0	47,6	n	n	n	y
usps.com	68	y	10	75	62,3	47,6	n	n	n	n
nih.gov	101	n	8	62	47,6	47,6	n	n	n	n
ca.gov	124	n	8	96	52,7	47,6	n	n	n	n
ed.gov	141	y	8	62	47,6	26,6	n	n	n	n
noaa.gov	199	n	8	96	52,7	77,1	n	n	n	n
weather.gov	228	n	12	62	71,5	77,1	n	n	n	n
census.gov	246	n	12	62	71,5	47,6	n	n	n	y
ssa.gov	276	n	7	36	36,2	36,2	n	n	n	n
nasa.gov	342	n	12	62	71,5	79,0	n	n	n	y
Other Sites										
U. of Phoenix	873	y	8	62	47,6	36,2	n	y	y	n
Columbia	1350	n	6	36	31,0	31,0	n	y	n	n
Northwestern	4457	n	8	10	26,6	31,0	n	n	n	n
VA	558	n	9	96	59,3	52,7	n	n	n	n
USAJobs	590	y	8	96	52,7	52,7	n	n	y	n
TreasuryDirect	2421	n	8	70	49,0	47,6	n	n	n	y
Twitter	31	y	6	10	19,9	19,9	n	n	y	y

Table 6: The German website sample (GER 2016) comprising 67 websites. Traffic ranks according to Alexa.

Website	Traffic Rank	Account created?	Min. Length	Size Charset	Min. Strength	Accepts Ads?	Places Ads?	User Choice
Top Traffic Sites								
Google.de	1	y	8	10	26,6	y	n	y
Amazon.de	2	y	6	10	19,9	y	y	y
Facebook.com	3	y	6	10	19,9	y	n	y
Ebay.de	5	y	6	36	31,0	y	n	y
Wikipedia	7	y	1	10	3,3	n	n	y
Web.de	8	y	8	10	26,6	y	n	y
Ebay-kleinanzeigen.de	9	y	6	10	19,9	y	n	y
T-online.de	10	y	8	36	41,4	y	n	y
Gmx.net	11	y	8	10	26,6	y	n	y
Bild.de	13	y	6	10	19,9	y	n	y
Yahoo.com	14	y	9	10	29,9	y	n	y
Spiegel.de	15	y	8	62	47,6	y	n	y
Xhamster.com	17	y	4	10	13,3	y	n	y
Paypal.com	18	y	8	10	26,6	n	n	y
Focus.de	19	y	8	10	26,6	y	n	y
Live.com	20	y	8	36	41,4	y	n	y
High Traffic Sites								
Mytoys.de	101	y	5	36	25,8	y	y	y
vodafone.de	102	y	8	36	41,4	n	y	n
aol.com	103	y	8	10	26,6	y	n	y
zdf.de	104	y	1	10	3,3	n	n	y
netflix.com	105	y	4	10	13,3	n	n	y
duden.de	106	y	6	96	39,5	y	n	y
eventim.de	107	y	5	10	16,6	n	y	y
xvideos.com	109	y	8	26	37,6	y	n	y
bonprix.de	110	y	6	10	19,9	n	y	y
Medium Traffic Sites								
proxer.me	491	y	8	10	26,6	y	n	y
Auto-motor-und-sport.de	493	y	5	10	16,6	y	n	y
pcgames.de	494	y	8	36	41,4	y	n	y
etsy.com	495	y	6	10	19,9	n	y	y
netdoktor.de	496	y	1	10	3,3	y	n	y
opodo.de	497	y	7	36	36,2	y	y	y
clipfish.de	499	y	5	10	16,6	y	n	y
Banks								
Deutsche Bank	74	y	5	10	16.6	n	y	n
KfW	2017	y	8	62	47,6	n	y	n
NordLB	13002	y	5	10	16,6	n	n	n
Deutsche Postbank	37	y	5	10	16,6	n	y	n
Ing-diba	89	n	5	10	16,6	n	y	y
DKB	125	n	5	36	25,8	n	y	y
comdirect	132	y	7	36	36,2	n	y	y
Volkswagenbank	1816	y	8	10	26,6	n	y	y
Consorsbank	410	y	5	10	16,6	n	y	y
Large Universities								
FU Hagen	1728	n	8	62	47,6	n	y	n
LMU München	869	n	4	36	20,7	n	n	n
U Köln	979	n	6	36	31,0	n	n	n
Goethe U Frankfurt	1235	n	6	10	19,9	n	n	n
Ruhr U Bochum	4989	n	8	10	26,6	n	n	n
WWU Münster	957	n	8	62	47,6	n	y	n
RWTH Aachen	1031	n	6	10	19,9	n	n	n
U Hamburg	1278	n	8	62	47,6	n	n	n

Continued on next page

Table 6 – continued from previous page

Website	Traffic Rank	Account created?	Min. Length	Size Charset	Min. Strength	Accepts Ads?	Places Ads?	User Choice
U Duisburg-Essen	1139	n	8	62	47,6	n	n	n
FAU Erlangen-Nürnberg	2664	n	4	10	13,3	n	n	n
Universities with top CS departments								
RWTH Aachen	1031	n	6	10	19,9	n	n	n
U Augsburg	3355	n	8	36	41,4	n	n	n
Jacobs U Bremen	15549	n	6	10	19,9	n	y	n
U Magdeburg	3750	n	6	34	30,5	n	n	n
Hasso-Plattner-Inst. Potsdam	16678	y	1	10	3,3	n	n	n
U Bayreuth	2363	n	8	10	26,6	n	n	n
TU Darmstadt	1990	y	9	96	59,3	n	n	n
FAU Erlangen-Nürnberg	2210	n	4	10	13,3	n	n	n
U Konstanz	3656	n	8	36	41,4	n	n	n
U des Saarlandes Saarbrücken	2368	n	6	36	31,0	n	n	n
Government Sites								
bundestag.de	2101	n	8	62	47,6	n	n	n
arbeitsagentur.de	97	n	8	62	47,6	n	n	n
bundesregierung.de	3440	n	8	62	47,6	n	n	n
bund.de	436	n	8	62	47,6	n	n	n
destatis.de	2240	n	8	96	52,7	n	n	n
bayern.de	245	n	8	44	43,7	n	n	n
nrw.de	309	n	8	10	26,6	n	n	n
europa.eu	377	n	8	10	26,6	n	n	n

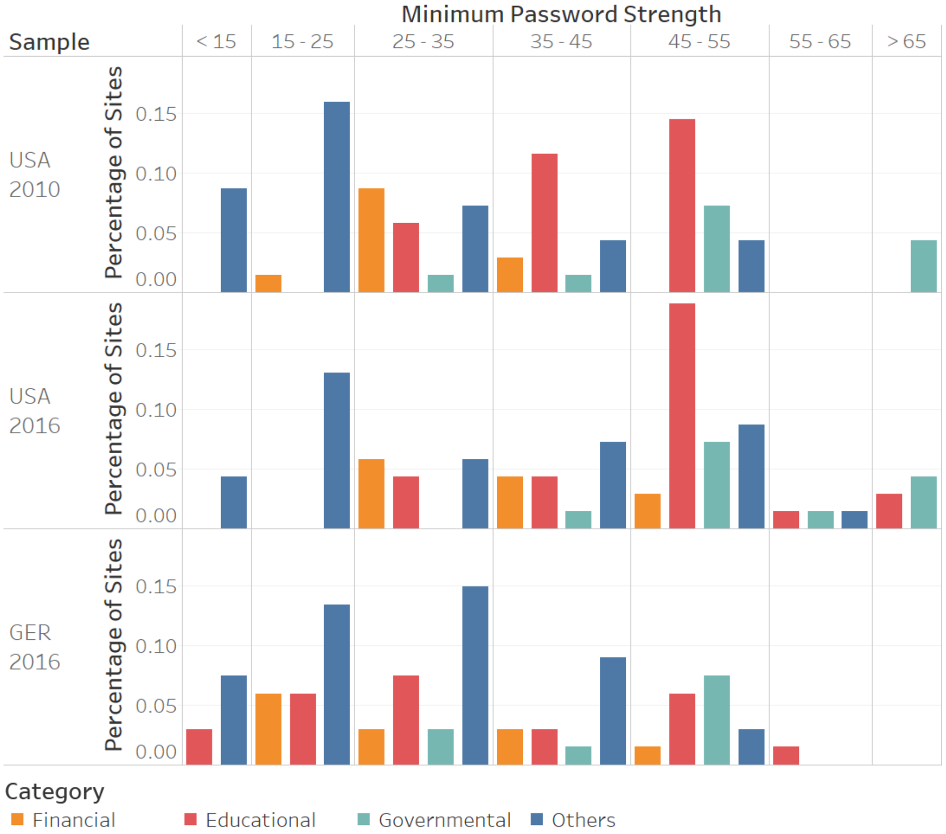

Figure 6: Histograms of the average PCP strength for all three samples along the different categories of websites.